Microbiota in Exhaled Breath Condensate and the Lung

Microbiota in Exhaled Breath Condensate and the Lung

Laura Glendinning[a] #, Steven Wright[a], Peter Tennant[a], Andrew C. Gill[a], David Collie[a], Gerry McLachlan[a]

The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, Midlothian, United Kingdom[a]

Running title: Evaluation of EBC for lung microbiota sampling

Address correspondence to Laura Glendinning, Laura.Glendinning@roslin.ed.ac.uk

Abstract word count: Abstract = 245 words, Importance = 140 words

Main text word count: 5225 words

Abstract

The lung microbiota is commonly sampled using relatively invasive bronchoscopic procedures. Exhaled breath condensate (EBC) collection potentially offers a less invasive alternative for lung microbiota sampling. We compared lung microbiota samples retrieved by protected specimen brushings (PSB) and exhaled breath condensate collection. We also sought to assess whether aerosolised antibiotic treatment would influence the lung microbiota and whether EBC was sensitive enough to detect such changes.

EBC was collected from 6 conscious sheep, and then from the same anaesthetised sheep during mechanical ventilation. Following the latter EBC collection, PSB samples were collected from separate sites within each sheep lung. On the subsequent day each sheep was then treated with nebulised colistimethate sodium. Two days after nebulisation, EBC and PSB samples were again collected. Bacterial DNA was quantified using 16S rRNA gene qPCR. The V2-V3 region of the 16S rRNA gene was amplified by PCR and sequenced using an Illumina Miseq. Quality control and operational taxonomic unit (OTU) clustering were performed within mothur.

EBC contained significantly less bacterial DNA than PSB samples. EBC samples from anaesthetised animals clustered separately by their bacterial community compositions in comparison to PSB samples and 37 bacterial OTUs were identified which were differentially abundant between the two sample types. Despite only low concentrations of colistin being detected in bronchoalveolar lavage fluid, PSB samples were found to differ by their bacterial compositions pre and post colistimethate sodium treatment. Our findings indicate that microbiota in EBC samples and PSB samples are not equivalent.

Importance

Sampling of the lung microbiota usually necessitates performing bronchoscopic procedures which involve a hospital visit for human participants and the use of trained staff. The inconvenience and perceived discomfort of participating in this kind of research may deter healthy volunteers and may not be a safe option for patients with advanced lung disease. This study set out to evaluate a less invasive method of collecting lung microbiota samples by comparing samples taken via protected specimen brushings (PSB) to those taken via exhaled breath condensate (EBC) collection. We found that there was less bacterial DNA in EBC samples compared to PSB samples and that there were differences between the bacterial communities in the two sample types. We conclude that while EBC and PSB samples do not produce equivalent microbiota samples, the study of the EBC microbiota may still be of interest.

47 **Introduction**

48 The study of the lung microbiota is a relatively new field in comparison to other areas of
49 microbiota research. Although an increasing number of studies are linking changes in the
50 composition of the lung bacterial communities to various disease states including allergies,
51 autoimmune disorders and inflammatory and infectious diseases (1), protocols for studying the
52 lung microbiota are not standardised, making comparisons between studies difficult.

53 One issue with studying the lung microbiota is the invasiveness of the sampling techniques; the
54 most common techniques are bronchoalveolar lavage (BAL) and the collection of protected
55 specimen brushings (PSB), both of which require the subject to undergo bronchoscopy. The
56 inconvenience and fear of complications associated with bronchoscopic procedures can lead to
57 healthy and/or diseased individuals declining to take part in studies (2), leading to a reduction in
58 the potential pool of volunteers for lung microbiota studies. It is also currently unknown whether
59 these sampling methods themselves can lead to changes in the lung microbiota.

60 Exhaled breath condensate (EBC) collection could potentially offer a less invasive method of
61 taking lung microbiota samples. This method involves condensing exhaled vapour into a liquid
62 and has previously been used to study exhaled bacteria, viruses and fungi (3-8). However, there
63 have been no studies using 16S rRNA gene sequencing to compare the bacteria found in EBC
64 samples to samples taken directly from the lungs. It is, therefore, not known whether it can be
65 used as a surrogate for more invasive sampling techniques. We sought to assess the feasibility of
66 using EBC in sheep to study the lung microbiota composition. We have previously used sheep as
67 a model for studying the lung microbiota (9, 10) due to the anatomical and immunological
68 similarity of their lungs to those of humans (11-13). In this study we compare EBC samples

collected from conscious sheep and from the same sheep under anaesthesia, to PSB samples taken from four spatially disparate sites within the lung.

We then extended this to address whether EBC analysis has the capacity to detect changes in bacterial community compositions by attempting to directly manipulate the lung microbiota with an inhaled antibiotic (colistimethate sodium (CMS): active against Gram-negative bacteria). During a previous study we examined the effect of intravenous CMS on the lung microbiota (9). Whilst we did identify changes in the lung microbiota composition, the longer term systemic antibiotic treatment used in this study is also likely to have affected the gut microbial populations. Immunological links between gut and lung immunity, the gut-lung axis, raise the possibility that such changes may have indirectly influenced the lung microbiota (14). In this study, we delivered nebulised CMS since this has been shown to lead to lower colistin plasma concentrations than injected CMS (15), enabling us to discern the direct effect of antibiotic treatment on respiratory bacterial communities.

A far greater quantity of bacterial DNA was isolated from PSB samples relative to EBC samples. We found that whilst there was some overlap between the types of bacteria found in these samples, EBC samples did cluster separately from PSB samples by their bacterial community compositions. Lastly, despite our antibiotic treatment regime only producing low concentrations of colistin in the lung epithelial lining fluids (the prodrug CMS is hydrolysed *in vivo* to the active form of the drug, colistin), significant differences in community composition were found between PSB samples derived pre- and post-treatment.

Applied and Environmental Microbiology

Methods and materials

Animals

Six commercially sourced, castrated, male Suffolk-cross sheep aged 14 months were used in this study. All animal experiments were approved by the Roslin Institute Animal Welfare and Ethics Committee and were subject to the Animals (Scientific Procedures) Act of 1986. Sheep had previously been housed outdoors as part of a large flock but were moved indoors before the study and remained indoors until the study end. Sheep were separated into two pens sharing the same airspace. One pen contained sheep ED951, ED952 and ED953 while the other contained sheep ED954, ED955 and ED956. The rectal temperatures and weights of all animals were taken prior to initial respiratory tract sampling. The animals weighed on average 49.2 ± 3.4 kg (mean ± standard deviation (SD)) and rectal temperatures were measured as 38.9 ± 0.89°C.

Experimental design

Conscious animals were confined in a yoke head restraint holding crate and EBC was collected for 10 mins using an RTubeVENT with cooling sleeve (Respiratory Research, Charlottesville, VA, USA) attached to a face mask. The sheep inhaled through a one-way inspiratory valve and expired through the RTubeVENT (**Figure 1**). The exhaled breath condensate samples from conscious sheep (EBC (cons)) were transferred from the RTubeVENT into Eppendorf tubes following the manufacturer's instructions and were frozen on dry ice within an hour of collection.

The sheep were then anaesthetised (3-5 hours later) following a procedure that has previously been described (16). Bronchoscopy was performed by use of an endotracheal tube. During

anaesthesia, EBC samples were collected for 10 mins by incorporating an RTubeVENT within the expiratory limb of the anaesthetic circuit (**Figure 1**). The condensate was again transferred into Eppendorf tubes. The exhaled breath condensate samples from the anaesthetised sheep (EBC (anaes)) were frozen on dry ice within an hour of collection. PSB samples (Disposable Microbiology Brush: ConMed, Utica, NY, USA) were taken from the left ventral diaphragmatic 1 (LVD1), right ventral diaphragmatic 1 (RVD1), right caudal diaphragmatic (RCD) and left caudal diaphragmatic (LCD) lung segments (**Figure 2**). Brushes were cut into phosphate buffered saline (PBS: D8537, Sigma-Aldrich, Irvine, UK) for storage. For each sampling day, an unused protected specimen brush was cut into PBS to act as a control.

Eighteen hours after recovery from anaesthesia, sheep were administered 2,000,000 IU of CMS in 4 ml distilled water by inhalation (Colomycin for Injection: Forest Laboratories UK ltd., Dartford, UK). Restraint of the conscious sheep was as described above and the CMS was delivered using a face mask connected via the inspiratory limb to an eflow rapid nebuliser (PARI Respiratory Equipment Inc., Midlothian, VA, USA). This treatment was repeated 6 hours later. Two days after the first CMS dose was administered, EBC (cons), EBC (anaes) and PSB samples were again collected as described above. Sheep were killed by barbiturate overdose and exsanguination and blood samples were collected. Blood was centrifuged at 2500g for 5 mins and the serum was removed and frozen on dry ice. Immediately post-mortem, 20 ml aliquots of PBS were used to collect BAL fluid. The urea concentration in plasma and BAL fluid was used to calculate the dilution factor of lung epithelial lining fluid in BAL fluid (17).

Quantitation of colistin in BAL fluid of sheep

133 BAL fluid was centrifuged at 1400g for 5 mins to remove cells prior to colistin quantification.
134 Quantitation of colistin in ovine BAL fluid essentially follows the method previously published
135 by Marchand et al. (18). Briefly, colistin sulfate (Item no. 17584 (mixture of A and B isoforms):
136 Cayman Chemicals, Ann Arbor, Ml, USA) was dissolved in H_2O to 1 mg/ml and a series of 7
137 calibrant solutions were created by diluting the stock solution into blank BAL fluid to cover the
138 range from 100 to 0.07 µg/ml. Polymyxin B (Sigma-Aldrich, Irvine, UK) was used as an internal
139 standard and was dissolved in water to 300 µg/ml. 2 µl of internal standard was added to 200 µl
140 of each of the calibrant solutions and to 200 µl of each of the test samples. 800 µl of H_2O &
141 0.1% (v/v) formic acid was added to each of the samples/calibrants and each was partially
142 purified by binding to a DSC-18 SPE cartridge (Sigma-Aldrich, Irvine, UK), eluting with 400 µl
143 MeOH & 0.1% (v/v) formic acid. The eluted fractions were dried under vacuum and
144 reconstituted in 50 µl of H_2O & 0.1% (v/v) formic acid for subsequent analysis.

145 All calibrants and samples were centrifuged at 13000g for 5 mins to pellet any precipitate and
146 then were analysed by online LC-MS/MS in duplicate. 5 µl aliquots were injected onto an Ace
147 Ultracore 2.5 SuperC18 HPLC column (75 x 2.1 mm) pre-equilibrated at 98% (v/v) buffer A,
148 where HPLC buffer A was H_2O with 0.1% (v/v) formic acid and 0.01% (v/v) trifluoroacetic acid,
149 whilst HPLC buffer B was acetonitrile with 0.1% (v/v) formic acid and 0.01% (v/v)
150 trifluoroacetic acid. The HPLC separation was developed by the following steps: from 2% buffer
151 B on 0 mins to 18% buffer B on 1 min; 22% buffer B on 3.5 mins; 100% buffer B on 4 mins;
152 100% buffer B on 5 mins and returning to 2% buffer B on 6 mins for 5 mins to re-equilibrate.
153 The flow rate was 200 µl/min and the eluent was passed directly to the electrospray source of an
154 Amazon ETD ion trap mass spectrometer (Bruker, Billerica, MA, USA) operated in positive ion
155 mode. The mass spectrometer was operated under multiple reaction monitoring conditions, using

parent ions of 578.3, 585.3 and 602.3 (representing the double charged ion of colistin B, A and polymyxin B respectively), fragmentation amplitudes of 0.8 and cut offs of 140 in each case. Calibration curves and colistin concentrations were calculated by Bruker's proprietary software QuantAnalysis, using the following reporter ions: colistin A 526.3, 535.3, 567.3, 576.3; colistin B 519.3, 528.3, 560.800, 569.3; Polymyxin B 543.300, 552.300, 584.300, 593.3.

DNA extraction

DNA extraction was carried out as described previously (10) using the MO BIO PowerSoil DNA Isolation Kit (MO BIO Laboratories inc., Carlsbad, CA, USA). All DNA extractions were carried out using the same lot of extraction kit, as the contamination present in different lots of the same make of kit have been shown not to be consistent (19). Samples were randomly assigned to one of four DNA extraction batches and for each of these batches an extraction kit reagent only control was produced (sample groupings can be found in **Dataset S1**).

16S rRNA gene amplification and sequencing

The V2-V3 variable regions of the 16S gene were amplified as described previously (10). A nested PCR protocol was used in order to decrease the potential bias introduced by the use of barcoded primers by only including primers with Illumina adaptor sequences and barcodes in the second PCR round (20). The first round used the V1-V4 primers 28F ('5-GAGTTTGATCNTGGCTCAG-3') and 805R ('5-GACTACCAGGGTATCTAATC-3') and the second round used the V2-V3 primers 104F ('5-GGCGVACGGGTGAGTAA-3') and 519R ('5-GTNTTACNGCGGCKGCTG-3') with Illumina adaptor sequences and barcodes (**Dataset S1**).

178 The PCR conditions for the first round were 94°C for 2 min followed by 20 cycles of 94°C for 1
179 min, 55°C for 45 s and 72°C for 1.5 min followed by 72°C for 20 min. The conditions for the
180 second round were 98°C for 30 s followed by 20 cycles of 98°C for 10 s, 67°C for 30 s and 72°C
181 for 10 s followed by 72°C for 2 min Q5 High-Fidelity 2X Master Mix (New England BioLabs.,
182 Ipswich, MA, USA) was used for all reactions. After each PCR round, amplicons were purified
183 using the AMPure XP PCR purification system (Beckman Coulter, Brea, CA, USA). The Human
184 Microbiome Project Mock Community HM-783D (obtained through BEI Resources, NIAID,
185 NIH) also underwent PCR alongside samples and controls. The Qubit dsDNA HS Assay Kit
186 (Thermo Fisher Scientific, Hemel Hempstead, UK) was used to calculate the quantity of DNA in
187 each sample then samples were pooled into a sequencing library. Sequencing was performed on
188 an Illumina Miseq (Illumina, San Diego, CA, USA) producing 250bp paired-end reads.

189

190 **Bioinformatic and statistical analysis**

191 Primers were removed with cutadapt (21) and sequences with greater than one base error per 10
192 bases were discarded. Quality control, taxonomic assignment and OTU clustering were
193 performed in mothur (22) as described previously (10). The data was subsampled to the
194 minimum number of sequence reads found in one of our samples (11675). Except where stated,
195 the following analyses were all performed within mothur.

196 Good's coverage values were calculated to estimate sample coverage (23). Distance matrices
197 were constructed using Yue Clayton theta values (24) and analysis of molecular variance
198 (AMOVA) was used to compare groups of samples by their bacterial composition (25).
199 Homogeneity of molecular variance (HOMOVA) was used to compare groups by their variation

(26). Principle coordinate analysis (PCOA) graphs were constructed to visualise sample clustering. The mothur command corr.axes was used to correlate bacterial OTUs to the axes of the PCOA graphs using the Spearman's rank correlation coefficient (*r*). The Bonferroni correction was used to correct for multiple statistical tests. The Inverse Simpson's index was employed to measure microbial diversity and the Chao 1 index was employed to measure richness. Metastats was used to identify OTUs which were significantly different between groups (27) except where more than two groups were compared in which case indicator analysis was used (28).

To compare groups statistically where data was non-parametric, the Mann-Whitney U test was used if the groups were independent and the Wilcoxon Signed Rank test was used where samples were related (performed in SPSS Statistics 21, IBM Analytics). Boxplots for qPCR data were constructed in SPSS. Heatmaps were constructed in R (Version 3.2.2, R Foundation for Statistical Computing (https://www.R-project.org)) using the packages gplots (29), heatplus (30), RColorBrewer (31) and Vegan (32).

qPCR

Quantification of the V3 region of the 16S rRNA gene was carried out using a previously described method (10). A standard curve was generated using DNA extracted from *Pseudomonas aeruginosa* strain PA0579 using 9 serial dilutions ranging from 14.2 ng/µl to 1.42 x 10^{-7} ng/µl (Quantified by the Qubit dsDNA HS Assay). The 0.142 ng/µl dilution served as a positive control for all qPCR reactions. The average CT value of no template controls was 28.7.

qPCR was performed using 1 µl of extracted DNA solution, the primers UniF340 ('5-ACTCCTACGGGAGGCAGCAGT-3') and UniR514 ('5-ATTACCGCGGCTGCTGGC-3') at a final concentration of 0.4 µM and the LightCycler 480 SYBR green I master mix (Roche Applied Science, Mannheim, Germany). The qPCR run consisted of a preincubation step (50°C for 2 min then 95°C for 10 s), an amplification step (45 cycles of 95°C for 30 s, then 63°C for 30 s) and a melting cycle.

Data availability

Sequencing reads can be accessed at Bioproject accession number PRJNA337937.

Results

Analysis of sequence quality and controls

DNA was extracted from respiratory samples and controls and the V2-V3 variable regions of the 16S rRNA gene were amplified by PCR, then sequenced. After forming contigs from forward and reverse reads, various quality controls steps were undertaken which reduced the total sequence numbers by 25.8%. The lowest Good's coverage estimate value amongst the samples was 0.996 indicating that at least 99.6% of the bacteria in this sample were identified. The sequence error rate was 0.18% and the average number of reads per sample was 39195 ± 11535 (mean ± SD). In total 867 OTUs were identified.

The Human Microbiome Project Mock Community HM-783D, containing the 16S rRNA genes of 20 bacterial species in staggered quantities and fixed ratios (1,000 to 1,000,000 copies per organism per µl), was processed alongside samples. Some biases were identified (**Dataset S2**). Three species were incorrectly identified to species level (*Acinetobacter baumannii* misidentified as *Acinetobacter rhizosphaerae*, *Clostridium beijerinckii* misidentified as *Clostridium butyricum* and *Neisseria meningitides* misidentified as *Neisseria cinerea*). Two of the bacterial species which were present in low numbers in the original community were not identified at any taxonomic level: *Actinomyces odontolyticus* and *Bacteroides vulgatus*. Their absence is likely due to the fact that they were in low abundance rather than the inability of our protocol to amplify and identify them, as they have previously been identified using the same protocol on a non-staggered version of the same mock community (10). We were also previously able to identify *Enterococcus faecalis* to genus level whereas in this study it could not be identified except potentially as OTU 10: Bacilli (class). This discrepancy, combined with the fact that

E.faecalis is in low abundance in the staggered mock community leads us to believe that this identification is incorrect.

As lung bacteria are in low abundance, lung samples are at particular risk from contamination by bacterial DNA originating from DNA extraction kit reagents. Therefore, as well as mock community controls, DNA extraction kit reagent controls were produced. DNA was extracted from samples in four batches and a reagent control was included with every batch. The bacterial OTUs identified in extraction kit controls did not occur consistently in samples from the same batch (**Figure 3**). Samples did cluster by DNA extraction batch (AMOVA: $P<0.001$) and 30 OTUs were found to be indicative of specific batches (**Dataset S3**). However, when these OTUs were removed from the dataset, samples still clustered by extraction batch (AMOVA: $P=0.014$), indicating that clustering was not entirely due to the presence of these OTUs. It is possible that some of these OTUs may be found naturally within the sheep respiratory system (eg. *Micrococcus luteus* – common coloniser of the human upper respiratory tract). We therefore decided not to remove these OTUs from our dataset. Since samples were randomly assigned to extraction batches, clustering by batch would be unlikely to lead to false positive statistical tests. However, there is the possibility that the presence of contaminating organisms may increase heterogeneity and thereby also increase stochastic noise.

In controls the most abundant OTUs on average were: PSB controls (*Corynebacterium*: 14.4%, Enterobacteriaceae: 10.9%, Intrasporangiaceae: 3.6%) and DNA extraction kit reagent controls (*Burkholderia*: 14.0%, Neisseriaceae: 10.5%, *Aggregatibacter*: 7.7%). The most abundant OTUs (on average) in the different sample types were: PSB samples (*Staphylococcus equorum*: 10.7%, *Mannheimia*: 6.5%, *Staphylococcus sciuri*: 5.6%), EBC (cons) (*Staphylococcus equorum*: 5.5%,

Neisseriaceae: 4.7%, *Paracoccus*: 4.3%) and EBC (anaes) (*Staphylococcus equorum*: 5.1%, *Staphylococcus epidermidis*: 3.7%, *Peptostreptococcus anaerobius*: 3.2%).

PSB samples contain more bacterial DNA than EBC samples

The V3 region of the 16S rRNA gene was quantified in our samples using qPCR. On average, PSB samples contained $1.53 \times 10^{-5} \pm 2.96 \times 10^{-5}$ ng/µl (mean ± SD) bacterial DNA (34200 ± 66100 16S copy numbers/µl) while EBC samples from conscious and anaesthetised sheep contained $4.28 \times 10^{-7} \pm 5.34 \times 10^{-7}$ ng/µl (955 ± 1190 16S copy numbers/µl) and $2.38 \times 10^{-7} \pm 7.12 \times 10^{-8}$ ng/µl (531 ± 159 16S copy numbers/µl) respectively (**Figure 4**). DNA extraction kit reagent only controls contained $1.82 \times 10^{-7} \pm 2.21 \times 10^{-8}$ ng/µl (406 ± 49 16S copy numbers/µl) while PSB controls and qPCR water controls contained $1.84 \times 10^{-7} \pm 1.05 \times 10^{-8}$ ng/µl (411 ± 23 16S copy numbers/µl) and $1.98 \times 10^{-7} \pm 2.06 \times 10^{-8}$ ng/µl (442 ± 46 16S copy numbers/µl) respectively.

All respiratory samples contained significantly more DNA than controls (Mann-Whitney: $P<0.005$ for all sample types). EBC samples from conscious and anaesthetised animals did not contain significantly different quantities of DNA (Wilcoxon signed-rank: P=0.182); however, PSB samples contained significantly more DNA when compared to both EBC (cons) (Wilcoxon signed-rank: P=0.002) and EBC (anaes) (Wilcoxon signed-rank: P=0.002) samples.

No significant clustering of EBC by sampling method

Since EBC samples from conscious sheep would be expected to include more bacteria from the upper respiratory tract than EBC samples from anaesthetised sheep, it was expected that these

two groups of samples would cluster separately from one another. However, no significantly separate clustering was observed (AMOVA: P=0.994). Despite this lack of separate clustering, EBC samples taken from the same sheep while it was conscious or anaesthetised did not contain the same bacterial communities, as can be observed in **Figure 5**.

The richness and diversity of bacterial communities were not significantly different between the two groups (Wilcoxon signed rank: P=0.583, P=0.595 respectively). When examined using metastats, there were significant differences in the quantities of several OTUs between these groups but all of these OTUs were present at low abundance (<1% abundant on average in each group).

PSB samples and EBC (anaes) samples cluster separately by their bacterial communities

We next investigated whether PSB and EBC samples contained equivalent bacterial communities. We compared PSB samples to EBC (anaes) samples as we hypothesised that these would be less likely to be contaminated by upper respiratory tract microbes than EBC (cons) samples. As well as containing a larger quantity of bacterial DNA, PSB samples also contained significantly different bacterial communities from EBC (anaes) samples (AMOVA: P=0.011, **Figure 6**). This may be explained by the difference in variation between the two groups (HOMOVA: P=0.026). Bacterial communities from PSB samples were also found to be significantly richer (Wilcoxon Signed Rank: P=0.006) but there was no significant difference in diversity (Wilcoxon Signed Rank: P=0.48). One OTU designated *Pseudomonas veronii*, which was the 4th most abundant OTU in PSB samples, was found to be significantly more abundant in PSB samples (Metastats q-value=0.046, PSB samples (mean ± SD): 3.9% ± 1.3%, EBC: only

318 one sequence read found in one sample). The *P.veronii* OTU was not found in any of the PSB
319 controls, indicating that it is likely that its presence is not due to contamination. This indicates
320 that the EBC samples do not simply contain a subset of the most abundant OTUs from PSB
321 samples. An additional 36 low abundance OTUs (<1% abundant on average in either group)
322 were found to be significantly different between the two groups by Metastats.

323 We considered that, since EBC (anaes) samples contained far less bacterial DNA than PSB
324 samples, they may have been more affected by contamination and this may be why these sample
325 types cluster separately. However, the five most abundant OTUs found in DNA extraction kit
326 reagent controls (*Burkholderia*, Neisseriaceae, *Aggregatibacter*, Pseudomonadaceae and
327 *Methylobacterium*) were not found to be significantly differently represented between PSB
328 samples and EBC (anaes) samples (Metastats q-value=1). It therefore seems unlikely that the
329 separate clustering of these groups is due merely to the increased effect of contamination on EBC
330 (anaes) samples.

331

332 **Changes in the bacterial communities found in respiratory samples pre and post CMS**
333 **treatment**

334 For both EBC (cons) and EBC (anaes) samples, pre and post treatment samples did not differ
335 significantly by bacterial community structure (AMOVA: P=0.449, P=0.094 respectively).
336 However, the bacterial communities found in PSB samples were found to be significantly
337 different pre and post treatment (AMOVA: P=0.014, F**igure 7**). This significantly separate
338 clustering was not merely due to differences in variation between the two groups (HOMOVA:
339 P=0.87). The OTU *P.veronii* was increased in post treatment samples (Metastats q-value:

340 P=0.043, Pre-treatment (mean ± SD): 0.74% ± 0.39%, Post-treatment: 7.1% ± 2.4%) and a
341 further 97 low abundance (<0.1%) OTUs were found to significantly differentiate pre and post
342 treatment samples.

343 Using the Wilcoxon signed rank test, it was found that the concentrations of DNA in respiratory
344 samples pre and post CMS treatment did not differ significantly: PSB samples (P=0.689), EBC
345 (cons) (P=0.345) and EBC (anaes) samples (P=0.248). The concentrations of colistin A
346 identified in sheep lungs are shown in **Table 1**.

Discussion

In this study we sought to identify whether invasive lung microbiota sampling techniques could be replaced by a less invasive method. We compared the quantity of bacterial DNA and the bacterial communities from samples taken by PSB and EBC collection in six sheep at two sampling points. EBC was collected from both conscious and anaesthetised animals. During mechanical ventilation the animals were intubated, meaning that the exhaled breath collected was derived only from the lower respiratory tract; by comparing these samples to those taken from conscious animals it should be possible to analyse the extent of contamination by bacteria from the upper respiratory tract on EBC (cons) samples. We found that EBC samples contained significantly less bacterial DNA than PSB samples and that PSB samples clustered separately by the composition of their bacterial communities to EBC (anaes) samples. EBC (anaes) and EBC (cons) samples did not cluster separately from one another.

Studies examining the utility of EBC for identifying lung colonising microorganisms have shown variable results. A study comparing EBC and sputum samples from asthma patients showed a 100% overlap in the culturable fungi identified between the two sample types (5) and a study examining the bacterial pathogens cultured from BAL and EBC samples in patients with ventilator associated pneumonia showed high concordance between the two sampling methods (33). In comparison, when PCR assays for ten common respiratory pathogens were performed on EBC and sputum samples from chronic obstructive pulmonary disease patients the results were found not to correlate well (34). EBC collection has also previously been found to be inefficient at detecting *Mycobacterium tuberculosis* (35), influenza viruses (36) and the common cystic fibrosis pathogens *Pseudomonas aeruginosa* and *Burkholderia cepacia* (37).

369 Some concerns have been raised about the use of EBC in respiratory research since the epithelial
370 lining fluid contained in these samples is often variable and is very highly diluted with water
371 vapour (38). This dilution could explain the far lower concentrations of bacterial DNA we
372 identified in EBC samples in comparison to PSB samples. It is also likely that PSBs would be
373 more efficient at sampling biofilms adhered to the lung mucosa which could explain some of the
374 differences observed between the two sample types. The difference between the bacterial
375 communities found in PSB samples and EBC may also be partially explained by how EBC is
376 formed. The exact origin of EBC is still under debate but it has been suggested that differences
377 observed between BAL and EBC samples could be explained by the fact that they sample
378 different compartments of the lung (39). Whilst it might be assumed that EBC would be derived
379 from both the central and peripheral airway compartments, which would perhaps explain the
380 differences between these samples and PSB samples, Bondesson et al. conclude that the majority
381 of EBC is in fact derived from the central airways (40). Without a better understanding of how
382 EBC is formed and what influences its composition, we are unable to account for the differences
383 we observed between the two sampling types.

384 Despite the fact that the concentrations of colistin found in the lungs were quite low after
385 nebulised CMS treatment, a significant difference was observed in the bacterial communities
386 from PSB samples pre and post treatment. In a previous study we found that the relative
387 proportion of Gram negative bacteria in the lung microbiota was reduced post injected CMS
388 treatment (excluding Pseudomonadales) (9). However, members of the Pseudomonadales
389 generally increased in relative abundance or remained stable after treatment. It is therefore
390 interesting to note that whilst in this study we did not find a significant reduction in the
391 abundance of Gram negative bacteria in PSB samples (data not shown), an OTU belonging to the

Pseudomonadales (*P.veronii*) was significantly increased in these samples post CMS treatment. It is possible that even at low concentrations the colistin may have had some effect on the lung bacteria or that the sampling strategy may itself in some way lead to changes in the lung microbiota, but at the moment this is purely speculative. All samples were randomised prior to DNA extraction and PCR amplification therefore observed differences were not due to samples from one time-point being processed separately from those from the other time-point.

In conclusion, the differences we observe between PSB samples and EBC samples lead us to not recommend using EBC collection as a replacement for more invasive lung sampling techniques. However, the EBC microbiota may still be an interesting avenue of study despite the fact that the small quantities of bacterial DNA in these samples does leave them more vulnerable to contamination and any future studies would have to be designed with this in mind.

403 **Funding information**

404 Funding for this study was supplied by the BBSRC (Biotechnology and Biological Sciences
405 Research Council, UK) under the EASTBIO doctoral training program (Grant number:
406 BB/J01446X/1) and, in part, by an Institute Strategic Programme Grant (BB/J004332/1) also
407 from the BBSRC. The funders had no input on study design; data collection and interpretation or
408 the decision to submit this study for publication.

Applied and Environmental Microbiology

AEM

Applied and Environmental Microbiology

409 **Acknowledgments**

410 The following reagent was obtained through BEI Resources, NIAID, NIH as part of the Human
411 Microbiome Project: Genomic DNA from Microbial Mock Community B (Staggered, Low
412 Concentration), v5.2L, for 16S rRNA Gene Sequencing, HM-783D. We would like to thank the
413 staff of Dryden Animal Services for the care of our animals.

References

1. **Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB.** 2016. The microbiome and the respiratory tract. Annu Rev Physiol **78:**481-504.
2. **Martinsen EMH, Leiten EO, Bakke PS, Eagan TML, Gronseth R.** 2016. Participation in research bronchoscopy: A literature review. Eur Respir J **3:**29511.
3. **Xu Z, Shen F, Li X, Wu Y, Chen Q, Jie X, Yao M.** 2012. Molecular and microscopic analysis of bacteria and viruses in exhaled breath collected using a simple impaction and condensing method. PLoS ONE **7:**e41137.
4. **Bhimji A, Singer LG, Kumar D, Humar A, Pavan R, Zhang H, Rotstein C, Keshavjee S, Mazzulli T, Husain S.** 2016. Feasibility of detecting fungal DNA in exhaled breath condensate by the Luminex Multiplex xTAG Fungal PCR Assay in lung transplant recipients: A pilot study. J Heart Lung Transplant **35:**S37.
5. **Carpagnano GE, Malerba M, Lacedonia D, Susca A, Logrieco A, Carone M, Cotugno G, Palmiotti GA, Foschino-Barbaro MP.** 2016. Analysis of the fungal microbiome in exhaled breath condensate of patients with asthma. Allergy Asthma Proc **37:**E41-E46.
6. **Carpagnano GE, Lacedonia D, Palladino GP, Logrieco G, Crisetti E, Susca A, Logrieco A, Foschino-Barbaro MP.** 2014. *Aspergillus* spp. colonization in exhaled breath condensate of lung cancer patients from Puglia Region of Italy. BMC Pulm Med **14:**22.
7. **Carpagnano GE, Lacedonia D, Cotugno G, Depalo A, Palmiotti GA, Giuffreda E, Foschino Barbaro MP.** 2015. Exhaled fungal microbiome in asthma: Possible association with severity and control. Eur Respir J **46:**PA4004.

8. **McDevitt JJ, Koutrakis P, Ferguson ST, Wolfson JM, Fabian MP, Martins M, Pantelic J, Milton DK.** 2013. Development and performance evaluation of an exhaled-breath bioaerosol collector for influenza virus. Aerosol Sci Tech **47:**444-451.

9. **Collie D, Glendinning L, Govan J, Wright S, Thornton E, Tennant P, Doherty C, McLachlan G.** 2015. Lung microbiota changes associated with chronic *Pseudomonas aeruginosa* lung infection and the impact of intravenous colistimethate sodium. PLoS ONE **10:**e0142097.

10. **Glendinning L, Wright S, Pollock J, Tennant P, Collie D, McLachlan G.** 2016. Variability of the sheep lung microbiota. Appl Environ Microbiol **82:**3225-3238.

11. **Bouljihad M, Leipold HW.** 1994. An ultrastructural study of pulmonary bronchiolar and alveolar epithelium in sheep. J Vet Med A **41:**573-586.

12. **Griebel PJ, Entrican G, Rocchi M, Beskorwayne T, Davis WC.** 2007. Cross-reactivity of mAbs to human CD antigens with sheep leukocytes. Vet Immunol Immunopathol **119:**115-122.

13. **Scheerlinck JPY, Snibson KJ, Bowles VM, Sutton P.** 2008. Biomedical applications of sheep models: From asthma to vaccines. Trends Biotechnol **26:**259-266.

14. **Dickson RP, Morris A.** 2017. Macrolides, inflammation and the lung microbiome: Untangling the web of causality. Thorax **72:**10-12.

15. **Yapa SWS, Li J, Patel K, Wilson JW, Dooley MJ, George J, Clark D, Poole S, Williams E, Porter CJH, Nation RL, McIntosh MP.** 2014. Pulmonary and systemic pharmacokinetics of inhaled and intravenous colistin methanesulfonate in cystic fibrosis patients: Targeting advantage of inhalational administration. Antimicrob Agents Chemother **58:**2570-2579.

16. **Collie D, Govan J, Wright S, Thornton E, Tennant P, Smith S, Doherty C, McLachlan G.** 2013. A lung segmental model of chronic *Pseudomonas* infection in sheep. PLoS ONE **8:**e67677.

17. **Rennard SI, Basset G, Lecossier D, Odonnell KM, Pinkston P, Martin PG, Crystal RG.** 1986. Estimation of volume of epithelial lining fluid recovered by lavage using urea as marker of dilution. J Appl Physiol **60:**532-538.

18. **Marchand S, Gobin P, Brillault J, Baptista S, Adier C, Olivier J-C, Mimoz O, Couet W.** 2010. Aerosol therapy with colistin methanesulfonate: A biopharmaceutical issue illustrated in rats. Antimicrob Agents Chemother **54:**3702-3707.

19. **Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW.** 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol **12:**87.

20. **Berry D, Ben Mahfoudh K, Wagner M, Loy A.** 2011. Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. Appl Environ Microbiol **77:**7846-7849.

21. **Martin M.** 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnetjournal **17:**10-12.

22. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF.** 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol **75:**7537-7541.

23. **Esty WW.** 1986. The efficiency of Good's nonparametric coverage estimator. Ann Stat **14:**1257-1260.

24. **Yue JC, Clayton MK.** 2005. A similarity measure based on species proportions. Commun Stat-Theory Methods **34:**2123-2131.

25. **Excoffier L, Smouse PE, Quattro JM.** 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics **131:**479-491.

26. **Stewart CN, Excoffier L.** 1996. Assessing population genetic structure and variability with RAPD data: Application to *Vaccinium macrocarpon* (American Cranberry). J Evol Biol **9:**153-171.

27. **Paulson J, Pop M, Bravo H.** 2011. Metastats: An improved statistical method for analysis of metagenomic data. Genome Biol **12:**P17.

28. **Dufrene M, Legendre P.** 1997. Species assemblages and indicator species: The need for a flexible asymmetrical approach. Ecol Monogr **67:**345-366.

29. **Warnes G, Bolker B, Bonebakker L, Gentleman R, Liaw W, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B.** 2015. gplots: Various R programming tools for plotting data, v2.17.0. http://CRAN.R-project.org/package=gplots.

30. **Ploner A.** 2015. Heatplus: Heatmaps with row and/or column covariates and colored clusters, v2.16.0. https://github.com/alexploner/Heatplus.

31. **Neuwirth E.** 2014. RColorBrewer: ColorBrewer palettes, v1.1-2. http://CRAN.R-project.org/package=RColorBrewer.

32. **Oksanen J, Blanchet F, Kindt R, Legendre P, Minchin P, O'Hara R, Simpson G, Solymos P, Stevens M, Wagner H.** 2016. Vegan: Community ecology package, v2.3-4/r2988. http://R-Forge.R-project.org/projects/vegan/}.

33. **May AK, Brady JS, Romano-Keeler J, Drake WP, Norris PR, Jenkins JM, Isaacs RJ, Boczko EM.** 2015. A pilot study of the noninvasive assessment of the lung

microbiota as a potential tool for the early diagnosis of ventilator-associated pneumonia. Chest **147:**1494-1502.

34. **Zakharkina T, Koczulla AR, Mardanova O, Hattesohl A, Bals R.** 2011. Detection of microorganisms in exhaled breath condensate during acute exacerbations of COPD. Respirology **16:**932-938.

35. **Jain R, Schriever CA, Danziger LH, Cho SH, Rubinstein I.** 2007. The IS6110 repetitive DNA element of *Mycobacterium tuberculosis* is not detected in exhaled breath condensate of patients with active pulmonary tuberculosis. Respiration **74:**329-333.

36. **St George K, Fuschino ME, Mokhiber K, Triner W, Spivack SD.** 2010. Exhaled breath condensate appears to be an unsuitable specimen type for the detection of influenza viruses with nucleic acid-based methods. J Virol Methods **163:**144-146.

37. **Vogelberg C, Hirsch T, Rosen-Wolff A, Kerkmann ML, Leupold W.** 2003. *Pseudomonas aeruginosa* and *Burkholderia cepacia* cannot be detected by PCR in the breath condensate of patients with cystic fibrosis. Pediatr Pulmonol **36:**348-352.

38. **Effros RM, Dunning MB, Biller J, Shaker R.** 2004. The promise and perils of exhaled breath condensates. Am J Physiol-Lung Cell Mol Physiol **287:**L1073-L1080.

39. **Jackson AS, Sandrini A, Campbell C, Chow S, Thomas PS, Yates DH.** 2007. Comparison of biomarkers in exhaled breath condensate and bronchoalveolar lavage. Am J Respir Crit Care Med **175:**222-227.

40. **Bondesson E, Jansson LT, Bengtsson T, Wollmer P.** 2009. Exhaled breath condensate-site and mechanisms of formation. J Breath Res **3:**016005.

531 **Figure Legends**

532 **Figure 1**: A) EBC was collected from conscious animals while they were restrained in a yoke
533 head restraint holding crate. A face mask was attached and sheep inhaled through a short tube
534 with an inlet valve and exhaled through an RTubeVENT. B) EBC was collected from
535 anaesthetised, mechanically ventilated animals by placing the RTubeVENT in-line with the
536 expiratory limb of the ventilator, near to the sheep's head.

537 **Figure 2**: Diagram of the sheep lung. PSB samples were taken pre and post colistimethate
538 sodium treatment from the RVD1 (right ventral diaphragmatic 1), LVD1 (left ventral
539 diaphragmatic 1), RCD (right caudal diaphragmatic) and LCD (left caudal diaphragmatic) lung
540 segments. Figure adapted with permission from previous work (16).

541 **Figure 3**: Heatmap showing samples grouped by batch, based upon the time DNA was extracted
542 from the samples. Bacterial OTUs were included where they had an abundance of ≥5% in at least
543 one sample. OTUs which were ≥5% abundant in a DNA extraction kit reagent control are
544 indicated by colour (batch 1 = green, batch 2 = pink, batch 3 = blue and batch 4 = yellow). DNA
545 extraction kit reagent controls are labelled as Extraction Kit Batch *n*. EBC samples from
546 conscious and from anaesthetised sheep are labelled EBC (cons) and EBC (anaes) respectively.
547 OTUs which were >5% abundant in an extraction kit control do not consistently appear in all
548 samples in the same batch.

549 **Figure 4**: Boxplot showing the log 16S rRNA gene concentrations found in sheep respiratory
550 samples (EBC samples from conscious and anaesthetised animals and PSB samples) and controls
551 (protected specimen brushes, DNA extraction kit reagents and qPCR reagents). Outliers were
552 defined by SPSS as either 'out' values (circles) or 'extreme' values (stars). PSB samples

553 contained significantly more bacterial DNA (P<0.005) than any other respiratory sample type or
554 control.

555 **Figure 5**: Heatmap showing EBC samples grouped by sheep and time-point. DNA extraction kit
556 reagent controls are labelled as Extraction Kit Batch *n*. EBC samples from conscious and from
557 anaesthetised sheep are labelled EBC (cons) and EBC (anaes) respectively. Bacterial OTUs were
558 included where they had an abundance of ≥5% in at least one sample. As can be observed, EBC
559 samples taken from the same sheep when it was conscious and when it was anaesthetised did not
560 necessarily contain the same bacterial OTUs.

561 **Figure 6**: PCOA graph showing the significantly separate clustering of EBC (anaes) and PSB
562 samples from sheep (AMOVA: P=0.011), which may be due to the difference in variation
563 between the two sample types (HOMOVA: P=0.026). The OTUs which most contributed to
564 samples moving in a positive or negative direction along either axis and which had p-values of
565 <0.00058 (defined by the Bonferroni correction as 0.5 divided by the total number of OTUs),
566 according to the corr.axes command within mothur, are listed. As this graph is only
567 representative of 20.3% of the total variability present between samples, caution should be taken
568 when interpreting how clustered sample groups appear.

569 **Figure 7**: PCOA graph showing the significantly separate clustering of PSB samples from sheep
570 pre and post CMS treatment (AMOVA: P=0.014). The OTUs which most contributed to samples
571 moving in a positive or negative direction along either axis and which had p-values of <0.00058
572 (defined by the Bonferroni correction as 0.5 divided by the total number of OTUs), according to
573 the corr.axes command within mothur, are listed. As this graph is only representative of 24% of
574 the total variability present between samples, caution should be taken when interpreting how
575 clustered sample groups appear.

576 **Tables**

577 **Table 1: Colistin A concentrations in sheep epithelial lining fluid**

Sheep	Colistin A* concentration (Mean ± SD)	Dilution of epithelial lining fluid in BAL	Colistin A concentration corrected for dilution (Mean)
ED951	0.346 ± 0.056 ng/µl	5.45	1.89 ng/µl
ED952	0.320 ± 0.034 ng/µl	4.18	1.34 ng/µl
ED953	0.290 ± 0.061 ng/µl	6.45	1.87 ng/µl
ED954	1.549 ± 0.251 ng/µl	15.75	24.40 ng/µl
ED955	0.625 ± 0.159 ng/µl	11.43	7.15 ng/µl
ED956	0.222 ± 0.017 ng/µl	29.5	6.56 ng/µl

578 *colistin B values were too low to be calculated accurately